Are Disagreements Honest?

Tyler Cowen
Robin Hanson*

Department of Economics
George Mason University

August 18, 2004 (First version April 16, 2001.)

*The authors wish to thank Curt Adams, Nick Bostrom, Geoff Brennan, James Buchanan, Bryan Caplan, Wei Dai, Hal Finney, Mark Grady, Patricia Greenspan, Kevin Grier, Robin Grier, Hans Haller, Claire Hill, Mary Hirschfeld, Dan Houser, Stephen Hsu, Michael Huemer, Maureen Kelley, Arnold Kling, Peter McCluskey, Tom Morrow, William Nelson, Mark Notturno, David Schmidtz, Susan Snyder, Aris Spanos, Damien Sullivan, Daniel Sutter, Alexander Tabarrok, William Talbott, Nicolaus Tideman, Eleizer Yudkowsky, and participants of the Virginia Tech economics department seminar for useful comments and discussion. We thank the Center for Study of Public Choice and the Mercatus Center for financial support.

* Correspondence to Robin Hanson, rhanson@gmu.edu, MSN 1D3, Carow Hall, Fairfax VA 22030-4444, 703-993-2326 FAX: 703-993-2323

Are Disagreements Honest?

ABSTRACT

We review literatures on agreeing to disagree and on the rationality of differing priors, in order to evaluate the honesty of typical disagreements. A robust result is that honest truth-seeking agents with common priors should not knowingly disagree. Typical disagreement seems explainable by a combination of random belief influences and by priors that tell each person that he reasons better than others. When criticizing others, however, people seem to uphold rationality standards that disapprove of such self-favoring priors. This suggests that typical disagreements are dishonest. We conclude by briefly considering how one might try to become more honest when disagreeing.

KEYWORDS: agreeing, disagree, common, prior, truth-seeking, Bayesian

I. Introduction

People disagree all of the time, especially about politics, morality, religion, and relative abilities. Virtually any two intelligent people can quickly find many topics of disagreement. Disagreements usually persist, and often become stronger, when people become mutually aware of them. Nor is disagreement usually embarrassing; it is often worse to be considered a "fence-sitter" without distinctive opinions.

Not only do people disagree; they often consider their disagreements to be about what is objectively true, rather than about how they each feel or use words. Furthermore, people often consider their disagreements to be honest, meaning that the disputants respect each other's relevant abilities, and consider each person's stated opinion to be his best estimate of the truth, given his information and effort.[1]

Yet according to well-known theory, such honest disagreement is impossible. Robert Aumann (1976) first developed general results about the irrationality of "agreeing to disagree." He showed that if two or more Bayesians would believe the same thing given the same information (i.e., have "common priors"), and if they are mutually aware of each other's opinions (i.e., have "common knowledge"), then those individuals cannot knowingly disagree. Merely knowing someone else's opinion provides a powerful summary of everything that person knows, powerful enough to eliminate any differences of opinion due to differing information.

Aumann's impossibility result required many strong assumptions, and so it seemed to have little empirical relevance. But further research has found that similar results hold when many of Aumann's assumptions are relaxed to be more empirically relevant. His results are robust because they are based on the simple idea that when seeking to estimate

[1] In this paper we consider only truth-seeking at the individual level, and do not attempt a formal definition, in the hope of avoiding the murky philosophical waters of "justified belief."

the truth, you should realize you might be wrong; others may well know things that you do not.

For example, this theory applies to any dispute that can be described in terms of possible worlds. This happens when people agree on what the answers would be in each imaginable world, but argue over which of these imaginable worlds is the real world. This theory can thus apply to disputes about facts that are specific or general, hard or easy to verify. It can cover the age of a car, the correctness of quantum mechanics, whether God created the universe, and which political candidate is more likely to induce prosperity. It can even apply to morality, when people believe there are objectively correct answers to moral questions.[2]

One of Aumann's assumptions, however, does make a big difference. This is the assumption of common priors, i.e., that agents with the same information must have the same beliefs. While some people do take the extreme position that that priors must be common to be rational, others take the opposite extreme position, that any possible prior is rational. In between these extremes are positions that say that while some kinds of priors and prior differences are rational, other kinds are not.

Are typical human disagreements rational? Unfortunately, to answer this question we would have to settle this controversial question of which prior differences are rational. So in this paper, we consider an easier question: are typical human disagreements honest? To consider this question, we do not need to know what sorts of differing priors are actually rational, but only what sorts of differences people seem to *think* are rational. If people mostly disagree because they systematically violate the rationality standards that

[2] The economics literature on disagreement is cited in more detail throughout the paper. Philosophers have long discussed disagreement, starting with Sextus Empiricus (2000, first edition predates 235 A.D.), who argued that when people disagree, they face an "equipollence" of reasons, and cannot judge their own perspective to be superior to that of others. See also Arnauld and Nicole, (1996 [1683]), Thomas Reid (1997 [1764]), Schiller (1934), Brandt (1944), Coady (1992), Nozick (1993), Rescher (1993), Everett (2001), and Goodin (2002).

they profess, and hold up for others, then we will say that their disagreements are dishonest.

After reviewing some stylized facts of disagreement, the basic theory of disagreement, how it has been generalized, and suggestions for the ways in which priors can rationally disagree, we will consider this key question of whether, in typically disagreements, people meet the standards of rationality that they seem to uphold. We will tentatively conclude that typical disagreements are best explained by postulating that people have self-favoring priors, even though they disapprove of such priors, and that self-deception usually prevents them from seeing this fact. We end by outlining some of the personal policy implications that would follow from this conclusion that human disagreement is typically dishonest.

II. The Phenomena of Disagreement

Before considering the theory of disagreement, let us now review some widely recognized stylized facts about human disagreement. Some of these "facts" may well be wrong, but it seems highly unlikely that most of them are wrong.

Virtually any two people capable of communication can quickly find a topic on which they substantially disagree. In such disagreements, both sides typically believe themselves to be truth-seekers, who honestly say what they believe and try to believe what is true. Both sides are typically well aware of their disagreement, and can reliably predict the direction of the other's next statement of opinion, relative to their own last statement.

Disagreements do not typically embarrass us. People are often embarrassed to discover that they have visibly violated a canon of rationality like logical consistency. Upon this discovery, they often (though not always) change their views to eliminate such violations. And in many cases (though again not always) fewer such violations tend to be discovered for individuals with higher IQ or superior training. Disagreements, however, happen

even though people are usually well aware of them, and high-IQ individuals seem no less likely to disagree than others. Not only are disagreements not embarrassing, more social shame often falls on those who agree too easily, and so lack "the courage of their convictions."

Real world disagreements seem especially frequent about relative abilities, such as who is smarter than whom, and about subjects, like politics, morality, and religion, where most people have strong emotional reactions. Discourse seems least likely to resolve disagreements of these kinds, and in fact people often move further away from each other's views, following a sustained dialogue.[3]

Psychologists suggest that human disagreements typically depend heavily on each person believing that he or she is better than others at overcoming undesirable influences on their beliefs (e.g., innuendo), even though people in fact tend to be more influenced than they realize (Wilson, Gilbert, & Wheatley 1998). Many people dismiss the arguments of others, often on the grounds that those others are less smart, knowledgeable, or otherwise less able. At the same time, however, such people do not typically accede to the opinions of those who are demonstrably equally or more able, be the relevant dimension IQ, life experience, or whatever. People are usually more eager to speak than they are to listen, the opposite of what a simple information-collection model of discussion would predict (Miller 2000).

The positions taken in many disagreements seem predictable, as in the saying that "where you stand depends on where you sit." In general, people seem inclined to believe what they "want to believe." For example, most people, especially men, estimate themselves to be more able than others and more able than they really are (Waldman 1998). Gilovich (1991, p.77) cites a survey of university professors, which found that 94% thought they were better at their jobs than their average colleagues. A survey of sociologists found that almost half said they expected to become among the top ten

[3] On the tendency for polarization, see Sunstein (1999).

leaders in the field (Westie 1973).[4] People also tend to think more highly of their groups, such as their home team's sporting success, their nation's military success, and their profession's social and moral value.

III. The Basic Theory of Agreeing to Disagree

Let us now consider the existing theory on disagreement. Most analysis of disagreement, like most analysis of inference and decision-making in general, has used Bayesian decision theory. Bayesian theory may not be fully satisfactory or fully general, but the core results in agreeing to disagree have been generalized beyond Bayesian agents to a greatf extent. Such generalizations have been possible because these core results rest mainly on a few simple intuitions. To see the intuitive appeal of the basic argument, consider the following simple parable.

Imagine that John hears a noise, looks out his window and sees a car speeding away. Mary also hears the same noise, looks out a nearby window, and sees the same car. If there was a shooting, or a hit-and-run accident, it might be important to identify the car as accurately as possible.

John and Mary's immediate impressions about the car will differ, due both to differences in what they saw and how they interpreted their sense impressions. John's first impression is that the car was an old tan Ford, and he tells Mary this. Mary's first impression is that the car was a newer brown Chevy, but she updates her beliefs upon hearing from John. Upon hearing Mary's opinion, John also updates his beliefs. They then continue back and forth, trading their opinions about the likelihood of various possible car features. (Note that they may also, but need not, trade evidence in support of those opinions.)

If Mary sees John as an honest truth-seeker who would believe the same things as Mary given the same information (below we consider this "common prior" assumption in

[4] For a survey of the psychology literature on this point, see Paulhus (1986).

detail), then Mary should treat John's differing opinion as indicating things that he knows but she does not. Mary should realize that they are both capable of mistaken first impressions. If her goal is to predict the truth, she has no good reason to give her own observation greater weight, simply because it was hers.

Of course, if Mary has 20/20 eyesight, while John is nearsighted, then Mary might reasonably give more weight to her own observation. But then John should give her observation greater weight as well. If they can agree on the relative weight to give their two observations, they can agree on their estimates regarding the car. Of course John and Mary might be unsure who has the better eyesight. But this is just another topic where they should want to combine their information, such as knowing who wears glasses, to form a common judgment.

If John and Mary repeatedly exchange their opinions with each other, their opinions should eventually stop changing, at which point they should become mutually aware (i.e., have "common knowledge") of their opinions (Geanakoplos and Polemarchakis 1982).[5] They will each know their opinions, know that they know those opinions, and so on.

We can now see how agreeing to disagree is problematic, given such mutual awareness. Consider the "common" set of all possible states of the world where John and Mary are mutually aware that John estimates the car age to be (i.e., has an "expected value" of) X, while Mary estimates it to be Y. John and Mary will typically each know many things, and so will know much more than just the fact that the real world is somewhere in this common set. But they do each know this fact, and so they can each consider, counterfactually, what their estimate would be if their information were reduced to just knowing this one fact. (Given the usual conception of information as sets of possible worlds, they would then each know only that they were somewhere in this common set of states.)

[5] For more on common knowledge, see Geanakoplos (1994), Bonnanno and Nehring (1999) and Feinberg (2000). For a critical view, see Koppl and Rosser (2002). For the related literature on "no-trade" theorems, see Milgrom and Stokey (1982).

Among the various possible states contained within the common set, the actual John may have very different reasons for his estimate of X. In some states he may believe that he had an especially clear view, while in others he may be especially confident in his knowledge of cars. But whatever the reason, everywhere in the common set John's estimate has the same value X. Thus if a counterfactual John knew only that he was somewhere in this common set, this John would know that he has some good reason to estimate X, even if he does not know exactly what that reason is. Thus counterfactual John's estimate should be X.

Similarly, if a counterfactual Mary knew only that she was somewhere in the common set, her estimate should be Y. But if counterfactual John and Mary each knew only that the real world is somewhere in this common set of possible worlds, they would each have exactly the same information, and thus should each have the same estimate of the age of the car. If John estimates the car to be five years old, then so should Mary. This is Aumann's (1976) original result, that mutual awareness of opinions requires identical opinions.[6]

The same argument applies to any dispute about a claim, such as whether the car is a Ford, which is true in some possible worlds and false in others. As long as disputants can imagine self-consistent possible worlds in which each side is right or wrong, and agree on what would be true in each world, then it should not matter whether the disputed claim is specific or general, hard or easy to verify, or about physical objects, politics, or morality.

A more detailed analysis says not only that people must ultimately agree, but also that the discussion path of their alternating expressed opinions must follow a random walk. Mary

[6] An argument for the irrationality of agreeing to disagree was also implicit in the classic "Dutch book" arguments for Bayesian rationality. These arguments showed that if an agent is willing to take bets on either side of any proposition, then to avoid guaranteed losses, his betting odds must satisfy the standard probability axioms. An analogous argument applies to a group of agents. If a group is to avoid combinations of bets that

should not be able to tell John how John's next opinion will differ from what Mary just said. Mary's best public estimate of John's next estimate must instead equal Mary's current best estimate (Hanson 2002).

Yet in ordinary practice, as well as in controlled laboratory experiments (Hanson and Nelson 2004), we know that disagreement is persistent. That is, people can and do consistently and publicly predict the direction of other people's opinion relative to their own opinion. For instance, if John first says the car is six years old, and Mary then says the car is three years old, a real Mary can usually accurately predict that John's next estimate will probably be more than three years. If Mary is rational, this suggests that John is not efficiently using the information contained in Mary's forecast.

IV. Generalizations of the Basic Theory

While Aumann's results depended on many strong assumptions, similar results obtain when these assumptions are considerably relaxed. For example, rather than knowing the exact values of each other's estimates, John and Mary need only be mutually aware of the fact that John thinks the car is at least as old as Mary thinks it is. (That is, a mutual awareness of the fact that X >= Y also implies that X=Y.) Larger groups of people need only identify the "extremist" among them, such as the person who has highest estimate (Hanson 1998). It is also enough for people to be mutually aware of a single summary statistic that increases whenever any person's estimate increases (McKelvey and Page 1986).

We also can relax the requirement that John and Mary be absolutely sure of the things they are mutually aware of, i.e., that they have "common knowledge." We need instead assume only "common belief." That is, we need only assume that there is some common set of possible states of the world where 1) some condition like X>=Y holds, and 2) both John and Mary believe that they are in this common set. John and Mary can sometimes

---

guarantee losses for the group as a whole, each group member must offer the same odds on every proposition.

be mistaken in this belief, but the higher their confidence, the smaller can be the difference between their estimates X and Y (Monderer and Samet 1989).

Thus John and Mary need not be absolutely sure that they are both honest, that they heard each other correctly, or that they interpret language the same way. Furthermore, if John and Mary each assign only a small chance to such confounding factors being present, then their difference of opinion must also be proportionately small. This is because while payoff asymmetries can induce non-linearities in actions, the linearity of probability ensures linearity in beliefs. A rational Mary's estimate of the car's age must be a linear weighted average of her estimate conditional on confounding factors being present, and her estimate conditional on the absence of such factors.

These results are also robust to John and Mary having many internal biases and irrationalities, as long as they also have a "rational core." Consider a corporation with many irrational employees, but with a truth-seeking Bayesian CEO in charge of its official statements. This CEO should treat inputs from subordinates as mere data, and try to correct for their biases. While such corrections would often be in error, this company's official statements would be rational, and hence would not agree to disagree with statements by other companies with similar CEOs. Similarly, if John and Mary were mutually aware of having "clear head" rational cores capable of suspecting bias in inputs from the rest of their minds, they should not disagree about the car.

We also need not assume that John and Mary know all logical truths. Through the use of "impossible possible states," Bayesians do not need to be logical omniscient (Hintikka 1975; Garber 1983). John and Mary (or their rational cores) do not even need to be perfect Bayesians, as similar results have been proven for various less-than-Bayesian agents (Rubinstein and Wolinsky 1990, Samet 1990, Geanakoplos 1994). For example, agents whose beliefs are represented by sets of probability distributions can be said to agree to disagree when they are mutually aware that their sets do not overlap (Levi 1974).

The beliefs of real people usually depend not only on their information about the problem at hand, but also on their mental context, such as their style of analysis, chosen assumptions, and recent thoughts. The existence of such features, however, is not by itself a reason to disagree. A truth-seeker who does not know which mental context is the most reliable should prefer to average over the estimates produced in many different mental contexts, instead of relying on just one random context.[7] So John should pay attention to Mary's opinion not only because it may embody information that John does not have, but also because it is the product of a different mental context, and John should want to average over as many mental contexts as he can.

This intuition can be formalized. Assume Mary has limited computational powers. Regardless of the computational strategies she uses, we can call Mary a "Bayesian wannabe" if she can imagine counterfactually being a Bayesian, and if she wants her actual estimates to be as close as possible to the estimates she would have if she were a Bayesian. It turns out that Bayesian wannabes who make a few simple calculations, and who would not agree to disagree about a state-independent variable, cannot agree to disagree about any matter of fact (Hanson 2003). Private information is irrelevant to estimating state-independent variables, since they taken on exactly the same value in every possible state.

V. Comparing Theory and Phenomena

The stylized facts of human disagreement are in conflict with the above theory of disagreement. People disagree, yet this theory says they should not. How can we resolve this conflict?

The theory above implicitly assumed that people say what they believe. Do people instead usually lie and not honestly state their opinions? Unfortunately for this

[7] John may have information suggesting that his mental context is better than random, but Mary may also have information on this topic. Persistent disagreement on this meta-topic should be no less problematic.

hypothesis, people usually have the strong impression that they are not lying, and it hard to see how people could be so mistaken about this. While there is certainly some element of sport in debates, and some recognition that people often exaggerate their views for effect, most people feel that they believe most of what they say when they disagree. People sometimes accuse their opponents of insincerity, but rarely accept this same label as a self-description. Even when they are conscious of steering a conversation away from contrary evidence, people typically perceive that they honestly believe the claims they make.

Another possibility is that most people simply do not understand the theory of disagreement. The arguments summarized above are complex in various ways, after all, and recently elaborated. If this is the problem, then just spreading the word about the theory of disagreement should eliminate most disagreement. One would then predict a radical change in the character of human discourse in the coming decades. The reactions so far of people who have learned about the theory of disagreement, however, do not lend much support to this scenario. Not only do such people continue to disagree frequently, it seems hard to find any pair of them who, if put in contact, could not frequently identify many persistent disagreements on matters of fact.

While Aumann's result is robust to generalizing many of his assumptions, it is not robust to generalizing the assumption of common priors. Bayesians can easily disagree due to differing priors, regardless of whether or not they have differing information, mental contexts, or anything else. Does this allow typical human disagreement to be rational?

To answer this question, we would need to not only identify the prior differences that account for typical human disagreements, we would also have to decide if these prior differences are rational. And this last topic turns out to be very controversial. We will soon review some arguments on this topic, but we will review them in the service of a more modest goal: evaluating whether typical human disagreement is honest. To evaluate the honesty of disagreement, we do not need to know what sorts of differing priors are actually rational, but only what sorts of differences people think are rational. We will

call disagreements dishonest when they are primarily the result of disputants who systematically violate the rationality standards that they profess and hold up for others.

VI. Proposed Rationality Constraints On Priors

Before reviewing arguments on the rationality of differing priors, let us review the nature of a Bayesian prior and prior-based disagreement. In general, Bayesian agents can have beliefs not only about the world, but also about the beliefs of other agents, about other agent's beliefs about other agents, and so on. When modeling agents for some particular purpose, the usual practice is to collect a "universe" of all possible states of the world that any agent in the model considers, or suspects that another agent may consider, and so on.

It turns out that one can always translate such agent beliefs into a "prior" probability distribution for each agent (Aumann 1998, Gul 1998). An agent's prior describes the probability she would assign to each possible state if her information were reduced to knowing only that she was somewhere in that model's universe of states. Many dynamic models contain an early point in time before the agents acquire their differing private information. In such models, the prior is intended to describe each agent's actual beliefs at this earlier time. In models without such an early time, however, the prior is interpreted counterfactually, as describing what agents would believe if sufficiently ignorant.[8]

By the nature of a prior, no agent can be uncertain about any other agent's prior; priors are by definition common knowledge among Bayesians. Thus when priors differ, all agents know those differences, know that they all know them, and so on. So while agents with differing priors can agree to disagree, they must anticipate such disagreements. Not only are they mutually aware that they disagree, they are also mutually aware that their disagreement is not due to the private information they each

[8] Note that even when priors are interpreted counterfactually, they have as much standing to be considered "real" as any other construct used to explain or justify spoken human

hold, whether that be information on the topic or information on their reasoning abilities. Each Bayesian knows exactly what every other agent would estimate if they had his information, and knows that this difference fully explains their disagreements. Relative to this estimate, he cannot publicly predict the future beliefs of another agent.[9]

Differing priors can clearly explain some kinds of disagreements. But how different can rational priors be? One extreme position is that no differences are rational (Harsanyi 1983, Aumann 1998). The most common argument given for this common prior position is that differences in beliefs should depend only on differences in information. If John and Mary were witnesses to a crime, or jurors deciding guilt or innocence, it would be disturbing if their honest rational beliefs -- the best we might hope to obtain from them -- were influenced by personal characteristics unrelated to their information about the crime. They should usually have no good reason to believe that the non-informational inputs into their beliefs have superior predictive value over the non-informational inputs into the beliefs of others.

Another extreme position is that a prior is much like a utility function: an ex post reconstruction of what happens, rather than a real entity subject to independent scrutiny. According to this view, one prior is no more rational than another than one utility function is more rational than another.[10] If we think we are questioning a prior, we are confused; what we are questioning is not a prior, but some sort of evidence. In this view priors, and the disagreements they produce, are by definition unquestionable.[11]

---

opinions, such as information sets, or epistemic principles. All such constructs are intrinsically counterfactual.

[9] This follows trivially from (Hanson 2002).

[10] One can accept this premise and still argue that priors should be treated as common. Given a prior, information set, and utility function that predict an agent's choices, one can predict those choices as well with any other prior, as long as one makes matching changes to their state-dependent utility. So one can argue that it is a convention of our language to describe agent differences in terms of differing utilities and information, rather than differing priors.

[11] Some (Bernheim 1986, Morris 1995) argue that multiple equilibria provide a rationale for differing priors, since each equilibrium describes different priors over game actions. In each equilibrium, however, agents would agree on the prior.

In the vast majority of disputed topics, the available evidence does not pin down with absolute certainty what we should believe. A consequence of this is that if there are no constraints on which priors are rational, there are almost no constraints on which beliefs are rational. People who think that some beliefs are irrational are thus forced to impose constraints on what priors are rational.

For example, technically we can think of a person at different moments in time as different agents, and we can even think of the different mental modules within a person's mind specializing in different mental tasks as different agents (Fodor 1983). If each different mental module at a different time could rationally have arbitrarily differing priors, then almost any sequence of belief statements a person might make might count as rational. Those who think that some sequences of statements are irrational must thus impose limits on how much priors can differ for agents close in space and time.

For example, it is common to require that the different mental modules within a single person share the same prior. Since it is infeasible for mental modules to share more than a limited amount of information with each other, we understand that different mental modules will sometimes give conflicting answers due to failing to share relevant information. Conflicts due to differing priors, however, seem less tolerable.

As another example, it is common to require Bayesians to change their beliefs by conditioning when they learn (or forget) information. That is, consider an earlier "self" who is the immediate causal ancestor of a later "self" who has learned a new fact about the world. While these different selves are logically two different agents who can in principle have different preferences and beliefs, it is common to say that the beliefs of the later self should typically be equal to the beliefs of the earlier self, conditional on that new fact. This is equivalent to saying that these two selves should base their beliefs on the same prior.[12]

[12] For more on rationality constraints for a single individual over time, see Van Fraassen (1984), Christensen (2000), Hurley (1989), Goldstein (1985), and Gilboa (1997).

Why exactly should these two selves have the same prior? If it were because one self is the immediate causal ancestor of the other, then by transitivity all causal ancestors and descendants should have the same prior. And if the process of conception, connecting parents and children, were a relevant immediate causal relation, then since all humans share a common evolutionary ancestor, all humans would have to have a common prior.

Most people do not go this far, and think that even if rationality requires a person to maintain the same prior over his lifespan and across his mental modules, it is also rational for humans to have differing priors at conception. This view, however, runs into the problem that it seems hard to believe that people are endowed at conception with DNA encoding detailed context-specific opinions on most of the topics on which they can later disagree. While there does seem to be a genetic component to some general attitudes (Olson, Vernon, Harris, & Jang 2001), the number of topics on which most people are capable of having independent detailed opinions is far larger than the number of bits in human DNA. Thus environmental influences must make an enormous contribution to human beliefs.

Even so, people do seem to be endowed early on with a few beliefs that could plausibly explain most of their disagreements. In particular, people tend to believe that they are better informed and better at resisting cognitive biases that other people. Psychologists explain human belief formation as due to not only to general attitudes, information, and experiences, but also to various random features of how exactly each person is exposed to a topic. People are influenced by their mood, how a subject was first framed, what other beliefs were easily accessible then, and so on. These random initial influences on beliefs, when combined with the tendency of each person to think he reasons better, can easily produce an unending supply of independent persistent disagreements.

How rational are such disagreements? Imagine that John believes that he reasons better than Mary, independent of any evidence for such superiority, and that Mary similarly believes that she reasons better than John. Imagine further that John accepts self-flattery

as a general pattern of human behavior, and so accepts it applying to himself counterfactually. That is, John can imagine counterfactually that he might have been Mary, instead of being John, and John's prior says that if he had been Mary, instead of John, he would have believed that Mary reasons better than John. Mary similarly thinks that if she were John, she would think that John reasons better.

Such priors are consistent in the sense that what John thinks he would believe if he were Mary, is in fact what Mary believes, no matter who "is" Mary. These priors are also "common" in the sense that everyone agrees about what Mary will think, no matter who really "is" Mary. These priors are not, however, "common" in the sense required for the theory of disagreement. Are such differing priors rational?

One argument against the rationality of such priors is that they violate "indexical independence." A non-indexical description of the state of the world includes facts like John and Mary's height and IQ. An indexical description of the world, in addition, says things like whether the "I" that would ordinary say, "I am John," instead says, "I am Mary." Indexical descriptions and information are need to, for example, describe what someone doesn't know when they have amnesia.

Under our ordinary concepts of physical causation, we expect to be able to predict non-indexical features of the world using only a rich enough set of other non-indexical features. For example, how likely John is to be right in his current argument with Mary may depend on John and Mary's experience, IQ, and education, but given a rich enough set of such relevant features, we do not expect to get more predictive ability from indexical information about who really "is" Mary. While we can imagine certain hypothetical scenarios where such predictive ability might arise, such as when John has amnesia but still knows he is very smart, these scenarios do not seem appropriate for priors.

Indexical independence is the assumption that John should behave the same no matter who really "is" John, and similarly for Mary or anyone else (Hanson 2004). And this

assumption is clearly violated by John's prior when it says that if he is John, John reasons better than Mary, but that if he is Mary, then Mary reasons better than John.

Finally, some theorists use considerations of the causal origins of priors to argue that certain prior differences are irrational. If John and Mary have different priors, they should realize that some physical process produced that difference. And if that difference was produced randomly or arbitrarily, it is not clear that John and Mary should retain it. After all, if John realized that some sort of memory error had suddenly changed a belief he had held for years, he would probably want to fix that error (Talbott 1990). So why should he be any more accepting of random processes that produced his earliest beliefs?

These intuitions can be formalized. One can argue that a rational Mary should be able to form coherent, even if counterfactual, beliefs about the chance that nature would have assigned her a prior different from the one she was actually given. Such counterfactual beliefs can be described by a "pre-prior." One can argue that Mary's actual prior should be consistent with her pre-prior in the sense that her prior should be obtained from her pre-prior by updating on the fact that nature assigned her a particular prior. Even if John and Mary have different pre-priors, if Mary thinks that it was just as likely that nature would have switched the assignment of priors, so that John got Mary's prior and vice versa, then John and Mary's priors should be the same. The priors about some event, like the car being a certain age, should also be the same if John and Mary believe that the chance of getting each prior was independent of this event (Hanson 2001).[13]

In summary, prior-based disagreements should be fully anticipated, and there are many possible positions on the question of when differing priors are rational. Some say no differences are rational, while others say all differences are rational. Many require agents close in space and time to have the same prior, but allow priors to differ at conception. While DNA cannot encode all future opinions, it might well encode a belief that you

[13] These assumptions about the causal origins of priors are satisfied by standard models of genetic inheritance, which predict that siblings (and parents and children) have almost

reason better, which could produce endless disagreements when combined with random influences on beliefs. Such a prior violates indexical independence, however, the idea that Mary's behavior may depend on her IQ, but not on who really "is" Mary. Rational prior differences are also limited when your prior must be consistent with your beliefs about the causal origins of your prior.

VII. Commonly Upheld Rationality Standards

Most people have not directly declared a position on the subject of what kinds of prior differences are rational. Most people would have to exert great effort to even understand these positions. So how can we figure out which rationality positions people uphold?

People are often reluctant to criticize the opinions of others, such as their boss to his face. There are, however, many situations in which people feel much freer to criticize. And in these situations, people often complain not about specific opinions, but about unreasonable patterns of opinions, patterns that seem to indicate faulty thinking. Thus one window into the rationality standards that people uphold and profess is the criticisms they make of others.

For example, people who feel free to criticize consistently complain when they notice someone making a sequence of statements that is inconsistent or incoherent. They also complain when they notice that someone's opinion does not change in response to relevant information. These patterns of criticism suggest that people uphold rationality standards that prefer logical consistency, and that prefer common priors for the mental modules within a person, and for his selves at nearby times.

Perhaps even more frequently, people criticize others when their opinions appear to have self-serving biases. For example, consider those sociologists, half of who expect to become among the top ten leaders in their field. Consider a school administrator who

---

the same ex ante chance of getting any particular DNA, and that these chances are correlated with little else.

favors his son for a school award, or a judge who does not excuse himself from a case in which he has an interest. Consider a manager who assigns himself to make an important sales presentation, or who uses his own judgment in an important engineering decision, rather than relying on apparently more qualified subordinates.

In such cases, interested observers who feel free to criticize commonly complain about self-favoring beliefs. The complaint is usually not that people say things they do not believe, but rather that they honestly believe things that favor them, without having sufficient reasons for such beliefs. Though critics acknowledge that self-favoring belief is a natural tendency, such critics do not seem to endorse those beliefs as accurate or reliable. Critics warn others, for example, not to be overly influenced to share such biased beliefs.

These common criticisms suggest that most people implicitly uphold rationality standards that disapprove of self-favoring priors, such as priors that violate indexical independence. These criticisms also suggest that people in fact tend to form beliefs as if they had such priors. That is, people do seem to think they can reason substantially better than others, in the absence of much evidence favoring this conclusion. People thus seem to violate the rationality standards they uphold. And as we have mentioned, such tendencies seem capable of explaining a great deal of human disagreement.

VIII. Truth-Seeking and Self-Deception

If we typically accept a rationality standard that disapproves of self-favoring priors, then why do we violate this standard with such enthusiasm? While people are usually embarrassed to learn they have been logical inconsistent, and smarter people do this less often, disagreements rarely embarrass us, and smarter people disagree just as often as others.[14]

[14] It is perhaps unsurprising that most people do not always spend the effort required to completely overcome known biases. What may be more surprising is that people do not simply stop disagreeing, as this would seem to take relatively little effort.

Non-truth-seeking and self-deception offer two complementary explanations for this difference in behavior. First, believing in yourself can be more functional that believing in logical contradictions. Second, while it is hard to deny that you have stated a logical contradiction, once the contradiction is pointed out, it is much easier to deny that a disagreement is due to your having a self-favoring prior. (You can always blame the other guy.)

On truth-seeking, while unbiased beliefs may be closer to the truth, self-favoring beliefs can better serve other goals. The virtues of self-confidence and self-esteem are widely touted (Benabou and Tirole 2002). Parents who believe in their children care more for them, and the best salesmen believe in their product, whether it is good or bad. By thinking highly of himself, John may induce Mary to think more highly of John, making Mary more willing to associate with John.

Scientists with unreasonably optimistic beliefs about their research projects may work harder and thus better advance scientific knowledge (Everett 2001; Kitcher 1990). Instead of simply agreeing with some standard position, people can better show off their independence and intelligence by inventing original positions and defending them. In response to our informal queries, numerous academics have told us that trying to disagree less would feel dishonest, destroy their identity, make them less human, and risk paralyzing self-doubt.

Self-favoring priors can thus be “rational” in the sense of helping one to achieve familiar goals, even if they are not “rational” in the sense of helping one to achieve the best possible estimate of the true situation (Caplan 2000).

Regarding self-deception, people seem more likely to gain the benefits of biased beliefs if they do not believe that they are biased (Taylor 1989). For example, a salesman is more persuasive when thinks he likes his product because of its features, rather that the fact that it is his product. And people do seem to often be unaware that they think highly of

themselves because of their prior. If Mary asks John to explain his high opinion of himself, John will usually point to some objective evidence, such a project he did well on. In response, John's critics will complain that he has succumbed to wishful thinking and self-deception.

Even if John attempts at some conscious levels to be unbiased, at other levels his mental programs may systematically bias his beliefs in the service of other goals. Our mental programs may under-emphasize evidence that goes against favored ideas, and distract the critical mechanisms that make us so adept at noticing and complaining about biases in other people's opinions (Mele 2001). The great and widely recognized power of flattery clearly shows that people have an enormous and widely recognized capacity for self-deception.

Which of these two explanations, non-truth-seeking and self-deception, is more fundamental? One important clue comes from the fact that academics who accept the conclusion that disagreement is irrational still disagree, including among themselves. When forced to overcome their self-deception and confront the issue, people consistently choose to continue to disagree. This suggests that, while this is not a conclusion they prefer to dwell on, most people fundamentally accept not being a truth-seeker.

The story we have outlined so far, of a widely recognized tendency toward self-favoring beliefs in others, together with self-deception about this tendency in ourselves, is commonly told in psychology and philosophy. Evolutionary arguments have even been offered for why we might have evolved to be biased and self-deceived.[15]

[15] Many have considered the evolutionary origins of self-deception and excess confidence one's own abilities (Waldman 1994). For example, truth-seekers who find it hard to lie can benefit by changing their beliefs (Trivers 1985; Trivers 2000). On topics like politics or religion, which are widely discussed but which impose few direct penalties for mistaken beliefs, our distant ancestors may have mainly demonstrated their cleverness and knowledge by inventing original positions and defending them well (Miller 2000).

This story is also commonly told in literature. For example, the concluding dream in Fyodor Dostoevsky's (1994 [1866]) Crime and Punishment seems to describe disagreement as the original sin, from which arises all other sins. In contrast, the description of the Houyhnhnms in Jonathan Swift's (1962 [1726]) Gulliver's Travels can be considered a critique showing how creatures (intelligent horses in this case) that agree too much lose their "humanity."

Given this story's ubiquity, its innate plausibility, and the fact that it fits the stylized facts of disagreement reasonably well, let us now accept this story as a working hypothesis to explain most human disagreement, and consider its implications.

VIII. How Few Meta-Rationals?

We can call someone a truth-seeker if, given his information and level of effort on a topic, he chooses his beliefs to be as close as possible to the truth. A non-truth seeker will, in contrast, also put substantial weight on other goals when choosing his beliefs. Let us also call someone *meta-rational* if he is an honest truth-seeker who chooses his opinions as if he understands the basic theory of disagreement, and abides by the rationality standards that most people uphold, which seem to preclude self-favoring priors.

The theory of disagreement says that meta-rational people will not knowingly have self-favoring disagreements among themselves. They might have some honest disagreements, such as on values or on topics of fact where their DNA encodes relevant non-self-favoring attitudes. But they will not have dishonest disagreements, i.e., disagreements directly on their relative ability, or disagreements on other random topics caused by their faith in their own superior knowledge or reasoning ability.

Our working hypothesis for explaining the ubiquity of persistent disagreement is that people are not usually meta-rational. While several factors contribute to this situation, a sufficient cause that usually remains when other causes are removed is that people do not

typically seek only truth in their beliefs, not even in a persistent rational core. People tend to be hypocritical in have self-favoring priors, such as priors that violate indexical independence, even though they criticize others for such priors. And they are reluctant to admit this, either publicly or to themselves.

How many meta-rational people can there be? Even if the evidence is not consistent with most people being meta-rational, it seems consistent with there being exactly one meta-rational person. After all, in this case there never appears a pair of meta-rationals to agree with each other. So how many more meta-rationals are possible?

If meta-rational people were common, and able to distinguish one another, then we should see many pairs of people who have almost no dishonest disagreements with each other. In reality, however, it seems very hard to find any pair of people who, if put in contact, could not identify many persistent disagreements. While this is an admittedly difficult empirical determination to make, it suggests that there are either extremely few meta-rational people, or that they have virtually no way to distinguish each other.

Yet it seems that meta-rational people should be discernable via their conversation style.[16] We know that, on a topic where self-favoring opinions would be relevant, the sequence of alternating opinions between a pair of people who are mutually aware of both being meta-rational must follow a random walk. And we know that the opinion sequence between typical non-meta-rational humans is nothing of the sort. If, when responding to the opinions of someone else of uncertain type, a meta-rational person acts differently from an ordinary non-meta-rational person, then two meta-rational people should be able to discern one another via a long enough conversation. And once they discern one another, two meta-rational people should no longer have dishonest disagreements.

[16] Aaronson (2004) has shown that regardless of the topic or their initial opinions, any two Bayesians have less than a 10% chance of disagreeing by more than a 10% after

Since most people have extensive conversations with hundreds of people, many of whom they know very well, it seems that the fraction of people who are meta-rational must be very small. For example, given N people, a fraction f of whom are meta-rational, let each person participate in C conversations with random others that last long enough for two meta-rational people to discern each other. If so, there should be on average $f^2CN/2$ pairs who no longer disagree.

If, across the world, two billion people, one in ten thousand of who are meta-rational, have one hundred long conversations each, then we should see one thousand pairs of people with only honest disagreements. If, within academia, two million people, one in ten thousand of who are meta-rational, have one thousand long conversations each, we should see ten agreeing pairs of academics. And if meta-rational people had any other clues to discern each another, and preferred to talk with one another, there should be far more such pairs. Yet, with the possible exception of some cult-like or fan-like relationships, where there is an obvious alternative explanation for their agreement, we know of *no* such pairs of people who no longer disagree on topics where self-favoring opinions are relevant

We therefore conclude that unless meta-rationals simply cannot distinguish each other, only a tiny non-descript percentage of the population, or of academics, can be meta-rational. Either few people have truth-seeking rational cores, and those that do cannot be readily distinguished, or most people have such cores but they are in control infrequently and unpredictably. Worse, since it seems unlikely that the only signals of meta-rationality would be purely private signals, we each seem to have little grounds for confidence in our own meta-rationality, however much we would like to believe otherwise.

IX. Personal Policy Implications

---

exchanging about a thousand bits, and less than a 1% chance of disagreeing by more than a 1% after exchanging about a million bits.

Readers need not be concerned about the above conclusion if they have not accepted our empirical arguments, or if they are willing to embrace the rationality of self-favoring priors, and to forgo criticizing the beliefs of others caused by such priors. Let us assume, however, that you, the reader, are trying to be one of those rare meta-rational souls in the world, if indeed there are any. How guilty should you feel when you disagree on topics where self-favoring opinions are relevant?

If you and the people you disagree with completely ignored each other's opinions, then you might tend to be right more if you had greater intelligence and information. And if you were sure that you were meta-rational, the fact that most people were not might embolden you to disagree with them. But for a truth-seeker, the key question must be how sure you can be that you, at the moment, are substantially more likely to have a truth-seeking, in-control, rational core than the people you now disagree with. This is because if either of you have some substantial degree of meta-rationality, then your relative intelligence and information are largely irrelevant except as they may indicate which of you is more likely to be self-deceived about being meta-rational.

One approach would be to try to never assume that you are more meta-rational than anyone else. But this cannot mean that you should agree with everyone, because you simply cannot do so when other people disagree among themselves. Alternatively, you could adopt a "middle" opinion. There are, however, many ways to define middle, and people can disagree about which middle is best (Barns 1998). Not only are there disagreements on many topics, but there are also disagreements on how to best correct for one's limited meta-rationality.

Ideally we would want to construct a model of the process of individual self-deception, consistent with available data on behavior and opinion. We could then use such a model to take the observed distribution of opinion, and infer where lies the weight of evidence, and hence the best estimate of the truth.[17] A more limited, but perhaps more feasible,

[17] Ideally this model would also satisfy a reflexivity constraint: when applied to disputes about self-deception it should select itself as the best model of self-deception. If most

approach to relative meta-rationality is to seek observable signs that indicate when people are self-deceived about their meta-rationality on a particular topic. You might then try to disagree only with those who display such signs more strongly than you do.

For example, psychologists have found numerous correlates of self-deception. Self-deception is harder regarding one's overt behaviors, there is less self-deception in a galvanic skin response (as used in lie detector tests) than in speech, the right brain hemisphere tends to be more honest, evaluations of actions are less honest after those actions are chosen than before (Trivers 2000), self-deceivers have more self-esteem and less psychopathology, especially less depression (Paulhus 1986), and older children are better than younger ones at hiding their self-deception from others (Feldman & Custrini 1988). Each correlate implies a corresponding sign of self-deception.

Other commonly suggested signs of self-deception include idiocy, self-interest, emotional arousal, informality of analysis, an inability to articulate supporting arguments, an unwillingness to consider contrary arguments, and ignorance of standard mental biases. If verified by further research, each of these signs would offer clues for identifying other people as self-deceivers.

Of course, this is easier said than done. It is easy to see how self-deceiving people, seeking to justify their disagreements, might try to favor themselves over their opponents by emphasizing different signs of self-deception in different situations. So looking for signs of self-deception need not be an easier approach than trying to overcome disagreement directly by further discussion on the topic of the disagreement.

We therefore end on a cautionary note. While we have identified some considerations to keep in mind, were one trying to be one of those rare meta-rational souls, we have no general recipe for how to proceed. Perhaps recognizing the difficulty of this problem can at least make us a bit more wary of our own judgments when we disagree.

---

people reject the claim that most people are self-deceived about their meta-rationality, this approach becomes more difficult, though perhaps not impossible.

X. Conclusion

A literature started by Robert Aumann, and spanning several decades, has explored the finding that, on matters of fact, honest disagreement is problematic. We have reviewed this literature, and found Aumann's initial result to be robust to many permutations, though not to introducing rationally differing priors. We reviewed arguments about which prior differences are rational, and found that the controversy surrounding this topic makes it difficult to determine whether typical disagreements are rational.

We can, however, use the rationality standards that people seem to uphold to find out whether typical disagreements are honest, i.e., are in accord with the rationality standards people uphold. We have suggested that when criticizing the opinions of others, people seem to consistently disapprove of self-favoring priors, such as priors that violate indexical independence. Yet people also seem to consistently use such priors, though they are not inclined to admit this to themselves or others.

We have therefore hypothesized that most disagreement is due to most people not being meta-rational, i.e., honest truth-seekers who understand disagreement theory and abide by the rationality standards that most people uphold. We have suggested that this is at root due to people fundamentally not being truth-seeking. This in turn suggests that most disagreement is dishonest.

We presented crude calculations suggesting that very few people can have much ground for thinking themselves meta-rational. This fact need not cause much concern for those willing to embrace the rationality of self-favoring priors, and to forgo criticizing the beliefs of others resulting from such priors. It also need not concern those who reject our empirical arguments. Those who accept our empirical arguments and who aspire to meta-rationality, however, face more difficulties, which we have briefly outlined.

References

Aaronson, Scott. “The Complexity of Agreement.” Unpublished manuscript, University of California at Berkeley, 2004.

Arnauld, Antoine and Nicole, Pierre. *Logic of the Art of Thinking*. Cambridge: Cambridge University Press, 1996 [1683].

Aumann, Robert J. "Agreeing to Disagree." *The Annals of Statistics*, 1976, 4, 6, 1236-1239.

Aumann, Robert J. “Common Priors: A Reply to Gul.” *Econometrica*, July 1998, 66, 4, 929-938.

Barns, EC, “Probabilities and epistemic pluralism.” *The British Journal for the Philosophy of Science*, March 1998. 49,1, 31-47.

Benabou, Roland, and Tirole, Jean, “Self-Confidence and Personal Motivation.” *Quarterly Journal of Economics* 117(3) : 871-915, August 2002.

Bernheim, R. Douglas. "Axiomatic Characterizations of Rational Choice in Strategic Environments." *Scandinavian Journal of Economics*, 1986, 88, 3, 473-488.

Bonnanno, Giacomo and Nehring, Klaus. "How to Make Sense of the Common Prior Assumption Under Incomplete Information." *International Journal of Game Theory*, 1999, 28, 409-434.

Brandt, Richard B. "The Significance of Differences of Ethical Opinion for Ethical Rationalism." *Philosophy and Phenomenological Research*, June 1944, 4, 4, 469-495.

Caplan, Bryan. "Rational Irrationality: A Framework for the Neoclassical-Behavioral Debate”, Eastern Economic Journal 26(2): 191-211, Spring 2000.

Christensen, David. "Diachronic Coherence versus Epistemic Impartiality." *The Philosophical Review*, July 2000, 109, 3, 349-371.

Coady, C.A.J. *Testimony: A Philosophical Study*. Oxford: Clarendon Press, 1992.

Dostoevsky, Fyodor. *Crime and Punishment*. Barnes and Noble Books, New York, 1994 [1866].

Everett, Theodore, “The Rationality of Science and the Rationality of Faith” *Journal of Philosophy*, 2001, 19-42.

Feinberg, Yossi. "Characterizing Common Priors in the Form of Posteriors." *Journal of Economic Theory*, 2000, 91, 127-179.

Feldman, Robert, and Custrini, Robert, “Learning to Lie and Self-Deceive” in *Self-Deception: An Adaptive Mechanism?,* edited by Joan Lockard & Delroy Paulhaus, Prentice Hall, 1988.

Garber, Daniel, “Old Evidence and Logical Omniscience in Bayesian Decision Theory”, in *Testing Scientific Theories*, edited by John Earman, University of Minnesota Press, 99-131.

Geanakoplos, John. “Common Knowledge.” *Handbook of Game Theory*, volume 2, edited by R.J. Aumann and S. Hart. Elsevier Science, 1994, 1437-1496.

Geanakoplos, John D. and Polemarchakis, Heraklis M. "We Can't Disagree Forever." *Journal of Economic Theory*, 1982, 28, 192-200.

Gilboa, Itzhak, “A Comment on the Absent-Minded Driver Paradox” *Games and Economic Behavior* 20(1) : 25-30, July 1997.

Gilovich, Thomas. *How We Know What Isn't So*. New York: Macmillan, 1991.

Gul, Faruk. “A Comment on Aumann’s Bayesian View.” *Econometrica*, July 1998, 66, 4, 923-927.

Goldstein, Michael. “Temporal Coherence” *Bayesian Statistics* 2, edited by Jose Bernardo, Morris DeGroot, Dennis Lindley, and Adrian Smith, 1985.

Goodin, Robert. “The Paradox of Persisting Opposition” *Politics, Philosophy, Economics*, 1, 2002.

Hanson, Robin. “Consensus By Identifying Extremists.” *Theory and Decision* 44(3):293-301, 1998.

Hanson, Robin. “Uncommon Priors Require Origin Disputes.” Unpublished manuscript, George Mason University, 2001.

Hanson, Robin. “Disagreement is Unpredictable.” *Economics Letters*, 77(3):365-369, November 2002.

Hanson, Robin. “For Savvy Bayesian Wannabes, Are Disagreements Not About Information?” *Theory and Decision*, 54(2):105-123, March 2003.

Hanson, Robin. “Priors over Indexicals”, Unpublished manuscript, George Mason University, 2004.

Hanson, Robin and Nelson, William. "An Experimental Test of Agreeing to Disagree." Unpublished manuscript, George Mason University, 2004.

Harsanyi, John. “Bayesian Decision Theory, Subjective and Objective Probabilities, and Acceptance of Empirical Hypotheses” *Synthese* 57, 341-365, 1983.

Hintikka, J. “Impossible Possible Worlds Vindicated” *Journal of Philosophical Logic* 4, 475-484, 1975.

Hurley, Susan. *Natural Reasons: Personality and Polity*. New York: Oxford University Press, 1989.

Kitcher, Philip. “The Division of Cognitive Labor” *The Journal of Philosophy*, 87, 1, 5-22, 1990.

Koppl, Roger and Rosser, Jr., J. Barkley, “All That I Have to Say Has Already Crossed Your Mind” *Metroeconomica* 53(4):339-360, 2002.

Levi, Isaac. “On Indeterminate Probabilities” *Journal of Philosophy* 71, 391-418, 1974.

McKelvey, Richard D. and Page, Talbot. "Common Knowledge, Consensus, and Aggregate Information." *Econometrica*, January 1986, 54, 1, 109-127.

Mele, Alfred R. *Self-Deception Unmasked*. Princeton: Princeton University Press, 2001.

Milgrom, Paul and Stokey, Nancy. "Information, Trade, and Common Knowledge." *Journal of Economic Theory*, 1982, 26, 17-27.

Miller, Geoffrey. *The Mating Mind, How Sexual Choice Shaped the Evolution of Human Nature*, Random House, New York, 2000.

Monderer, Dov and Samet, Dov. "Approximating Common Knowledge with Common Beliefs." *Games and Economic Behavior*, 1989, 1, 170-90.

Morris, Stephen. “The Common Prior Assumption in Economic Theory.” *Economics and Philosophy*, 1995, 11, 227-253.

Olson, James, Vernon, Philip, Harris, Julie, and Jang, Kerry. “The Heritability of Attitudes: A Study of Twins” *Journal of Personality and Social Psychology*, June 2001, 80(6) 845-860.

Nozick, Robert. *The Nature of Rationality*. Princeton: Princeton University Press, 1993.

Paulhus, Delroy L. "Self-deception and Impression Management in Test Responses." In Angleitner, A. & Wiggins, J. S., *Personality assessment Via Questionnaires*. New York, NY: Springer, 1986, 143-165.

Reid, Thomas. *An Inquiry into the Human Mind*. University Park, Pennsylvania: Pennsylvania State University Press, 1993 [1764].

Rescher, Nicholas. *Pluralism: Against the Demand for Consensus*. Clarendon Press, Oxford, 1993.

Rubinstein, A., Wolinsky, A.(1990) "On the Logic of Agreeing to Disagree Type Results" *Journal of Economic Theory* 51 184 -193.

Samet, D. (1990). "Ignoring Ingorance and Agreeing to Disagree." *Journal of Economic Theory* 52, 190 -207.

Schiller, F.C.S. *Must Philosophers Disagree?* London: Macmillan and Co., Limited, 1934.

Sextus Empiricus. *Outlines of Scepticism*. Cambridge: Cambridge University Press, 2000, first edition predates 235 A.D.

Sunstein, Cass. "The Law of Group Polarization." University of Chicago Law School, John M. Olin Law & Economics Working Paper, no.91, December 1999.

Swift, Jonathan. *Gulliver's Travels and Other Writings*, New York: Bantam Books, 1962 [1726].

Talbott, William J. *The Reliability of the Cognitive Mechanism, A Mechanistic Account of Empirical Justification*, 1990, Garland Publishing, New York.

Taylor, S.E. *Positive Illusions: Creative Self-Deception and the Healthy Mind.* New York: Basic Books, 1989.

Trivers, Robert, *Social Evolution*, Benjamin/Cummings, Menlo Park, Ca., 1985.

Trivers, Robert, "The Elements of a Scientific Theory of Self-Deception," in *Evolutionary Perspectives on Human Reproductive Behavior*, Annals of the New York Academy of Sciences, edited by Dori LeCroy and Peter Moller, volume 907, April 2000.

Van Fraasen, C. "Belief and the Will." *Journal of Philosophy*, May 1984, 81, 5, 235-256.

Waldman, Michael. "Systematic Errors and the Theory of Natural Selection." *American Economic Review*, June 1994, 84, 3, 482-497.

Westie, Frank R. "Academic Expectations of Professional Immortality: A Study of Legitimation." *The American Sociologists*, February 1973, 8, 19-32.